

Improved Approximation for Weighted Tree Augmentation with Bounded Costs

David Adjiashvili

Institute for Operations Research, ETH Zürich
Rämistrasse 101, 8092 Zürich, Switzerland
email: addavid@ethz.ch

Abstract

The Weighted Tree Augmentation Problem (WTAP) is a fundamental well-studied problem in the field of network design. Given an undirected tree $G = (V, E)$, an additional set of edges $L \subseteq V \times V$ disjoint from E called *links*, and a cost vector $c \in \mathbb{R}_{\geq 0}^L$, WTAP asks to find a minimum-cost set $F \subseteq L$ with the property that $(V, E \cup F)$ is 2-edge connected. The special case where $c_\ell = 1$ for all $\ell \in L$ is called the Tree Augmentation Problem (TAP). Both problems are known to be NP-hard.

For the class of bounded cost vectors, we present a first improved approximation algorithm for WTAP since more than three decades. Concretely, for any $M \in \mathbb{Z}_{\geq 1}$ and $\epsilon > 0$, we present an LP based $(\delta + \epsilon)$ -approximation for WTAP restricted to cost vectors c in $[1, M]^L$ for $\delta \approx 1.96417$. For the special case of TAP we improve this factor to $\frac{5}{3} + \epsilon$.

Our results rely on a new LP, that significantly differs from existing LPs achieving improved bounds for TAP. We round a fractional solution in two phases. The first phase uses the fractional solution to decompose the tree and its fractional solution into so-called β -simple pairs losing only an ϵ -factor in the objective function. We then show how to use the additional constraints in our LP combined with the β -simple structure to round a fractional solution in each part of the decomposition.

1 Introduction

The Weighted Tree Augmentation Problem (WTAP) is a well-studied problem in the field of network design. Given an undirected tree $G = (V, E)$, an additional set of edges $L \subseteq V \times V$ disjoint from E called *links*, and a cost vector $c \in \mathbb{R}_{\geq 0}^L$, WTAP asks to find a minimum-cost set $F \subseteq L$ with the property that $(V, E \cup F)$ is 2-edge connected. Recall that a graph is 2-edge connected if there are at least 2 edge-disjoint paths between any two nodes. The special case where $c_\ell = 1$ for all $\ell \in L$ is called the Tree Augmentation Problem (TAP).

WTAP is widely recognized as one of the fundamental problems in the field of network design (see e.g. the surveys of Kuhler [9] and Kortsarz and Nutov [11]). The main open problem is whether there is an approximation algorithm for WTAP with factor better than 2 (we review the literature in the next section). Our main result is an improved approximation algorithm for the case of bounded costs.

Theorem 1. *Let $\delta = \frac{8(23+3\sqrt{5})}{121} \approx 1.96418$. For any fixed $M \in \mathbb{Z}_{\geq 1}$ and $\epsilon \in \mathbb{R}_{>0}$, there exists an LP-based polynomial-time $(\delta + \epsilon)$ -approximation algorithm for WTAP restricted to instances with cost vectors satisfying $c \in [1, M]^L$, i.e. whenever $1 \leq c_\ell \leq M$ holds for all $\ell \in L$.*

For TAP we obtain an improved approximation guarantee as stated hereafter.

Theorem 2. *For any fixed $\epsilon \in \mathbb{R}_{>0}$, there exists an LP-based polynomial $(\frac{5}{3} + \epsilon)$ -approximation algorithm for TAP.*

Both algorithms achieve approximations with respect to a new linear programming (LP) relaxation, which we call *the bundle LP*. To the best of our knowledge, our result for WTAP is the first approximation algorithm that achieves a factor better than 2 for all trees, and any non-trivial family of cost functions. Furthermore, both bounds are the best known among those that are based on an LP.

Our results are based on several new ideas combined with classical results for WTAP, which we briefly explain next. First, we explain the well-known set covering reformulation of WTAP. We denote by $V[H]$ and $E[H]$ the node set and the edge set of a graph H , respectively. Associate with every link $\ell = uv \in L$ the unique path $P_{uv} \subseteq E[G]$ in G connecting u and v . Now, it is easy to verify that a set of links $S \subseteq L$ is a feasible solution for the WTAP instance at hand, if and only if the union of the corresponding paths covers the edge set of G , namely if $\cup_{\ell \in S} P_\ell = E[G]$.

For a set $X \subseteq E[G]$ denote by $\text{cov}(X) \subseteq L$ the set of links that cover at least one edge of X . When $X = \{e\}$ is a singleton we write $\text{cov}(e)$ instead of $\text{cov}(\{e\})$. The *natural LP relaxation* for WTAP, also known as *the cut LP*, is based on the latter reformulation. It contains one variable x_ℓ for each link $\ell \in L$ and asks to solve

$$\text{minimize } \sum_{\ell \in L} c_\ell x_\ell \quad \text{subject to} \quad (1)$$

$$\sum_{\ell \in \text{cov}(e)} x_\ell \geq 1 \quad \forall e \in E[G], \quad (2)$$

$$x_\ell \geq 0 \quad \forall \ell \in L. \quad (3)$$

The bundle LP, the optimal solution of which we round to obtain our results, is defined as follows. For an integer $\gamma \in \mathbb{Z}_{\geq 1}$, a γ -*bundle* is a union of γ (not necessarily distinct) paths in G . See Figure 1 for an example. Denote by \mathcal{B}_γ the set of all γ -bundles in G . In words, for a carefully chosen γ , the bundle LP contains, on top of the constraints from the natural LP, constraints that ensure that each γ -bundle is covered in the fractional solution by links with sufficiently high cost. Formally, the LP contains the constraints

$$\sum_{\ell \in \text{cov}(B)} c_\ell x_\ell \geq \text{OPT}(B) \quad \forall B \in \mathcal{B}_\gamma, \quad (4)$$

where for any $X \subseteq E[G]$, $\text{OPT}(X) \in \mathbb{R}_{\geq 0}$ is the minimum cost of a set of links in L that covers all edges in X . The latter constraints are clearly valid for any integral solution, as any such

2 is tight for them, even in the case of TAP. The factor 2 has since only been improved for special classes of trees, including a $(1 + \ln 2)$ -approximation for the case of bounded-diameter trees by Cohen and Nutov [4]. Improving the factor 2 for WTAP is a major open problem in network design [9, 11]. To the best of our knowledge, our algorithm for WTAP is the first improvement in the approximation guarantee (over the factor 2) for any special case of WTAP and all trees for over three decades.

In contrast, several approximation algorithms with factors better than 2 are known for TAP. The first such algorithm, achieving a factor $1.875 + \epsilon$ was given by Nagamochi [15]. This was later improved to 1.8 by Even et al. [5]. The current best algorithm is a $\frac{3}{2}$ -approximation, due to Kortsarz and Nutov [13]. An improved $\frac{17}{12}$ -approximation was developed by Maduel and Nutov [14] for the special case of TAP, where each link connects two leaves of the tree.

The techniques used to achieve the latter improved algorithms for TAP are combinatorial in nature, and seem very hard to modify for an improved approximation of WTAP. In an effort to improve the approximation factor for WTAP, several algorithm with approximation factors better than 2 have recently been developed for TAP, that are based on continuous relaxations of the problem. Along these lines, Kortsarz and Nutov [12] recently showed an LP-based $\frac{7}{4}$ -approximation algorithm. Our $(\frac{5}{3} + \epsilon)$ -approximation for TAP improves on that factor for LP-based algorithms. Cheriyan and Gao [1] presented an approximation with respect to a semidefinite program (SDP) obtained from Lasserre tightening of an LP relaxation with factor $\frac{3}{2} + \epsilon$. A strong point of both papers compared to our result is that the mathematical program is used in the analysis, while the algorithms themselves are combinatorial.

The bundle LP differs from all existing LPs that were recently proved to have integrality gap better than 2. Indeed, most existing such LPs include, on top of the constraints of the natural LP, constraints that exploit structural properties of feasible, or optimal solutions restricted to very specific structures, such as *twin links*, *stems* etc. (see [12] for formal definitions of these and other related notions). In contrast, as we discussed before, the bundle LP attempts to uniformly cut off fractional solutions that have low cost due to *all* sufficiently simple obstructions which have large integrality gap.

Another important open problem is the integrality gap of the natural LP relaxation. An upper bound of 2 on the integrality gap of the cut LP follows from some of the aforementioned 2-approximations, including that of Jain [8] and Goemans et al. [7]. There is also a lower bound of $\frac{3}{2}$ known on the integrality gap due to Cheriyan et al. [3]. The true integrality gap, however, still remains open, even for the special case of TAP.

WTAP can also be interpreted as a problem of covering a laminar family with point-to-point links (see e.g. [12]). This observation implies that the problem of augmenting a k -edge connected graph to a $(k + 1)$ -edge connected graph can be reduced to WTAP, whenever k is odd, due to the laminar structure of the family of minimum cuts, in this case.

2 Preliminaries

We delay certain proofs of technical results to Appendix B, and the treatment of the improved rounding algorithm for TAP to Appendix C.

Our algorithm relies on a few basic results for WTAP and some related problems that we recap here. For a finite set N and a vector $x \in \mathbb{R}_{\geq 0}^N$, let $\text{supp}(x) = \{i \in N \mid x_i > 0\}$ denote the support of x . For a subset $Y \subseteq N$ denote $x(Y) = \sum_{i \in Y} x_i$. For an integer $k \in \mathbb{Z}_{\geq 1}$ denote $[k] = \{1, \dots, k\}$. To distinguish between edges of the tree and links, we write $e = \{u, v\}$ and $\ell = uv$ for an edge e connecting u and v and a link ℓ connecting u and v , respectively.

Throughout the paper we use the notion of *contraction* of edges and links. By contracting an edge $e = \{u, v\} \in E[G]$ we mean replacing the nodes u and v by a new *compound node* w and connecting every edge $e' \in E[G]$ different from e that had a connection to either u , or v to the new node w . We also update the set of links in an analogous way: A link that was connected to either u , or v is connected after contracting e to w . If the link was connected to both u and v it becomes a self-loop of w . By contracting a set of edges $F \subseteq E$ we mean contracting the

edges in F one after the other in any order. By contracting a link (or a set of links) we mean contracting the set of edges of the tree covered by the link (or set of links). Figure 2 illustrates the contraction operation.

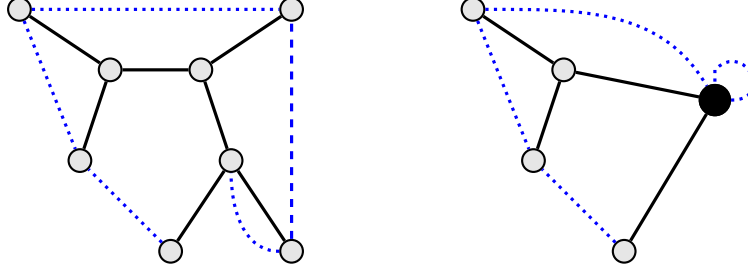


Figure 2: An illustration of a contraction operation. The links are shown with dotted and dashed lines. The dashed link in the left instance is contracted to obtain the instance on the right. The full black node is the obtained compound node

We refer to the book of Schrijver [16] for further details and results on some of the notions presented in this section. Furthermore, throughout the paper we do not optimize the running time of algorithms to facilitate a cleaner presentation. Next, we introduce the well-known notions of shadows of links and shadow completeness.

Shadows and shadow completion. For a tree G and a link $\ell = uv \in V[G] \times V[G]$ denote by $P_\ell^G \subseteq E[G]$ the u - v path in G . When G is clear from the context we drop the superscript and write P_ℓ . Let $\ell, \ell' \in V[G] \times V[G]$ be two links. We say that ℓ' is a *shadow* of ℓ if $P_{\ell'} \subseteq P_\ell$. An instance (G, L) is *shadow complete* if for every $\ell \in L$ all shadows of ℓ are also in L . We can always assume that the instance is shadow complete: If it is not, then all shadows of links in L can be added to L . A cost of an added link ℓ is the minimum cost of any original link in L , of which ℓ is a shadow. The latter shadow completing operation does not change the optimal value of the instance. Furthermore, shadows that are added in the completion can always be replaced in any solution by original links that cover a superset of the edges covered by the shadow without increasing the cost of the solution. We will hence always assume that the instance is shadow complete.

Up-links and a simple LP-based 2-approximation. One important building block that is used throughout the algorithm is a simple 2-approximation algorithm for WTAP that rounds a fractional solution $x \in \mathbb{R}_{\geq 0}^L$ for the natural LP relaxation. As we mentioned before, the latter is achievable with several existing algorithms. We present a simple such algorithm here for completeness and since it outlines ideas that we use later on.

The algorithm starts by rooting the tree G at an arbitrary node $r \in V[G]$. For every link $\ell = uv \in L$ denote by $\text{nca}(uv) \in V[G]$ the nearest common ancestor of u and v in G . We call a link $\ell = uv$ an *up-link* if $\text{nca}(uv) \in \{u, v\}$. Denote by $L^{up} \subseteq L$ the set of all up-links. The following simple rounding lemma is driving the approximation algorithm.

Lemma 3. *Let $x \in \mathbb{R}_{\geq 0}^L$ be a feasible solution of the natural LP on instance (G, L, c) with the property $\text{supp}(x) \subseteq L^{up}$. Then there is feasible set of links $S \subseteq L^{up}$ with cost $c(S) \leq c^\top x$, that can be computed in polynomial time.*

The 2-approximation algorithm can easily be obtained from Lemma 3 as follows. For every $\ell = uv \in L \setminus L^{up}$ with $\ell \in \text{supp}(x)$ set $w = \text{nca}(uv)$ and adapt the solution x to obtain a solution x' as follows. Set $x'_{uw} = x_{uw} + x_\ell$, $x'_{vw} = x_{vw} + x_\ell$, $x'_\ell = 0$ and $x'_{\ell'} = x_{\ell'}$ for all other links ℓ' . This eliminates ℓ from the support of x by incurring an additional cost of $c_\ell x_\ell$, without breaking the feasibility of the fractional solution. By repeating this for all links in the support that are

not up-links we obtain a feasible solution with cost at most $2c^\top x$, as desired (see Figure 3 for an illustration). This transformation and Lemma 3 imply the following proposition.

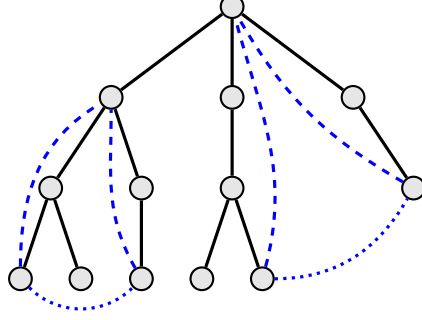


Figure 3: An illustration of up-links, and the simple 2-approximation algorithm. Two regular links (dotted lines) and their corresponding pairs of up-links (dashed lines).

Proposition 4. *There is a polynomial algorithm that given an instance (G, L, c) of WTAP and a fractional solution $x \in \mathbb{R}_{\geq 0}^L$ to the natural LP returns a solution $S \subseteq L$ for the WTAP instance with cost $c(S) \leq 2c^\top x$.*

Star-shaped WTAP and edge-covers. Given a graph G , an *edge cover* is a subset $S \subseteq E[G]$ of the edges of the graph that is incident to every node of G , i.e. such that for every $u \in V[G]$, there exists an edge $e \in S$ such that $u \in e$. A *fractional edge cover* is a vector $x \in \mathbb{R}_{\geq 0}^{E[G]}$ with the property that $x(\{e \in E[G] \mid u \in e\}) \geq 1$ for every $u \in V[G]$.

The minimum-cost edge cover problem is the problem of finding minimum-cost set of edges that is an edge cover of G . It is well-known that the problem is solvable in polynomial time. More importantly for our purposes, the problem can be used to model a special case of WTAP, which we define next.

Definition 5. An instance (G, L, c) of WTAP is called *star-shaped* if there exists a node $r \in V[G]$ such that r is incident to P_ℓ for every link $\ell \in L$. A node satisfying the latter condition is called a *hub*.

The following property of star-shaped instances is important to establish the connection to edge covers. We call an edge in a tree a *leaf edge* if it is incident to a leaf in the tree.

Lemma 6. *A solution $S \subseteq L$ for a star-shaped instance (G, L, c) of WTAP is feasible if and only if S covers all leaf edges.*

Lemma 6 directly implies that in star-shaped instances, all links that are not incident to any leaf can be removed from L , as they are redundant in any feasible solution. We hence assume that every link in L touches at least one leaf.

Lemma 6 also implies that WTAP with star-shaped instances (G, L, c) is equivalent to the edge cover problem: Construct a graph H whose node set is $V[H] = U \cup \{r\}$, where U is the set of leaf in G , and r is any hub. The set of edges $E[H]$ contains one edge e^ℓ for every link $\ell \in L$ with cost c_ℓ . If ℓ connects two leaves $u, v \in U$, set $e^\ell = \{u, v\}$. Otherwise, if ℓ connects a leaf $u \in U$ to an internal node, set $e^\ell = \{u, r\}$. Finally, we add a single self-loop e_0 connected to r with cost zero. This self-loop removes the need to incur cost for covering the node r .

It is immediate that edge covers in H and feasible solutions to WTAP instance (G, L, c) are in one-to-one correspondence: $C \subseteq E[H]$ is an edge cover if and only if $\{\ell \in L \mid e^\ell \in C\}$ is a feasible WTAP solution. Figure 4 illustrates the latter transformation.

The most important implication from the latter correspondence is that one can round a fractional solution to the natural LP relaxation of a star-shaped WTAP instance with a loss

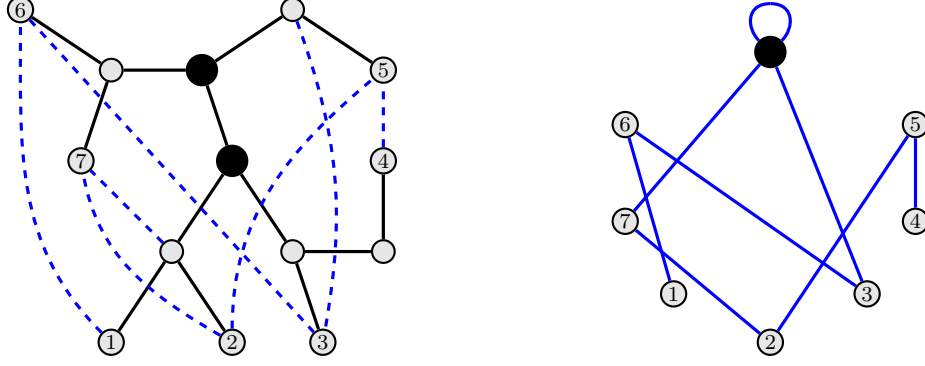


Figure 4: An illustration of the transformation into an edge cover problem. Left: A star shaped instance with hubs indicated as full nodes. Right: The corresponding edge cover instance.

of factor $\frac{4}{3}$. The following classical result on the integrality gap of the fractional edge cover polytope and the latter transformation imply this result, which we summarize in Proposition 8.

Lemma 7. *There is a polynomial algorithm that given a graph G , a cost vector $c \in \mathbb{R}_{\geq 0}^{E[G]}$ and a fractional edge cover $x \in \mathbb{R}_{\geq 0}^{E[G]}$ computes an edge cover $C \subseteq E[G]$ with cost $c(C) \leq \frac{4}{3}c^\top x$.*

Proposition 8. *There is a polynomial algorithm that given a star-shaped WTAP instance (G, L, c) and a fractional solution $x \in \mathbb{R}_{\geq 0}^L$ of the natural LP for this instance computes a feasible solution $S \subseteq L$ for the WTAP instance with cost $c(S) \leq \frac{4}{3}c^\top x$.*

3 A Rounding Algorithm for WTAP (Proof of Theorem 1)

We present next an algorithm that rounds an optimal solution $x \in \mathbb{R}_{\geq 0}^L$ to LP_γ , for a constant γ that depends on the accuracy ϵ and the maximum weight M . We determine γ explicitly later. To this end we outline the general strategy.

We round the solution in two phases. In the first phase we break the tree G into a union of simpler subtrees, and equip each subtree with a feasible fractional WTAP solutions, derived from x by using a simple splitting operation. Each fractional solution will only use links connecting nodes in its corresponding subtree, allowing us to treat each subtree and its corresponding solution separately. To bound the cost of the decomposition it is first necessary to guarantee that no edge in the tree is over-covered, i.e. covered in x by a fraction larger than some constant $a = a(\epsilon)$. This is achieved by rounding an appropriately scaled version of the fractional solution x , using Proposition 4, and contracting the obtained set of links.

In the second phase the fractional solutions in each subtree are rounded to integral solutions using two different procedures. Here, an important structural property of each subtree-solution pair in the decomposition is used, which is called *simplicity*, and is parametrized by an integer β . The links of each β -simple pair are partitioned into two types. Then, depending on which type of links dominates the cost in the fractional solution, one of the two rounding procedures is used to obtain an integral solution, as each rounding procedure achieves a good approximation with respect to one type of links. The bundle constraints are exploited in one of the rounding procedures, while the other rounding procedure uses the fact that instances corresponding to β -simple pairs are close to being star-shaped. Finally, the union of all solutions from all subtrees in the decomposition is returned as the solution.

3.1 Phase One: Decomposition

The decomposition of G and x is obtained in two steps, which we describe hereafter.

3.1.1 Contraction of Heavily Covered Edges

In the first step we select a low-cost set of links to cover all edges of G that are covered by a total weight of at least some constant $a = a(\epsilon)$. Concretely, define the set of edges that are *heavily covered* as

$$E^h = \left\{ e \in E \mid x(\text{cov}(e)) \geq \frac{2}{\epsilon} \right\},$$

namely E^h are the edges that are covered by x with links with a mass of at least $\frac{2}{\epsilon}$. To obtain the desired set of links we contract the edges $E \setminus E^h$ to obtain the subtree G^h of G whose edge set is E^h . Now, the solution x covers *every* edge in G^h by a fraction of at least $\frac{2}{\epsilon}$. It follows that $y = \frac{\epsilon}{2} \cdot x$ is a feasible solution to the natural LP relaxation of the WTAP instance on G^h , and hence Proposition 4 can be applied. The result is a set of links $L_0 \subseteq L$ that cover all edges of G^h and has cost $c(L_0) \leq 2c^\top y = \epsilon c^\top x$, i.e. its cost is only an ϵ -fraction of the cost of x . The links L_0 are included in the solution, so it henceforth remains to cover all edges in $E[G] \setminus E^h$.

Let \bar{G} be the tree obtained by contracting the edges in E^h . Note that $E[\bar{G}] = E[G] \setminus E^h$, so the edges of \bar{G} are exactly the ones that we still need to cover. The key property of the solution x , interpreted in the new tree \bar{G} , is the following *thin coverage property*, which states that for every $e \in E[\bar{G}]$ it holds that

$$x(\text{cov}(e)) \leq \frac{2}{\epsilon} = O\left(\frac{1}{\epsilon}\right),$$

a property that we crucially exploit to bound the cost of the decomposition step.

We note that x might not be a solution to LP_γ on \bar{G} , for the same γ that we used to obtain x as a solution for G . This is due to the fact that γ -bundles in \bar{G} might contain compound nodes, obtained by contracting some links in L_0 , and hence they might not represent γ -bundles of G . However, as we will later show, x maintains enough of the structure given by the γ -bundle constraints to bound the integrality gap. To make these arguments precise later on we keep track of which nodes of the new tree \bar{G} are compound nodes, namely nodes that represent more than one original node of the tree G , and were created by contracting some links in L_0 . These nodes are denoted by V^{cp} and called *early compound nodes* to distinguish them from compound nodes obtained due to later contractions that we perform in the algorithm. Furthermore, for $u \in V^{cp}$, we denote by $s_u \in \mathbb{Z}_{\geq 0}$ the total cost of links of L_0 that were contracted to obtain the early compound node u . For non-compound nodes $u \in V[\bar{G}] \setminus V^{cp}$ we set $s_u = 0$. Since links cover paths in G , one link of L_0 cannot contribute to the formation of more than one early compound node, so we have

$$\sum_{u \in V[\bar{G}]} s_u = \sum_{u \in V^{cp}} s_u = c(L_0).$$

3.1.2 Decomposition

In the next step the algorithm decomposes the instance into simpler instances, by breaking the tree at certain edges using a simple splitting operation. Each part of the obtained decomposition is a pair (T, z) , where T is a subtree of \bar{G} , and z is a fractional solution for the WTAP instance restricted to T . The decomposition is obtained by an iterative greedy procedure that employs the following operation, that we call *splitting*.

Definition 9. Let (G, L, c) be a WTAP instance and let $z \in \mathbb{R}_{\geq 0}^L$. Let $e = \{u, v\} \in E$ be any edge. Let G^u and G^v be the trees obtained by removing e from G , where G^u is the tree that contains u . The *splitting of z at e* produces two vectors $z^u \in \mathbb{R}_{\geq 0}^L$ and $z^v \in \mathbb{R}_{\geq 0}^L$ defined as follows. We define z^u ; z^v is defined symmetrically. For $\ell = pq \in L$ set

$$z_\ell^u = \begin{cases} z_\ell & \text{if } p, q \in V[G^u] \setminus \{u\} \\ 0 & \text{if } \{p, q\} \cap V[G^v] \neq \emptyset \\ z_\ell + \sum_{\ell' \in \text{cov}(e), q \in \ell'} z_{\ell'} & \text{if } p = u, q \in V[G^u]. \end{cases}$$

Note that $\text{supp}(z^u) \subseteq V[G^u] \times V[G^u]$ and $\text{supp}(z^v) \subseteq V[G^v] \times V[G^v]$.

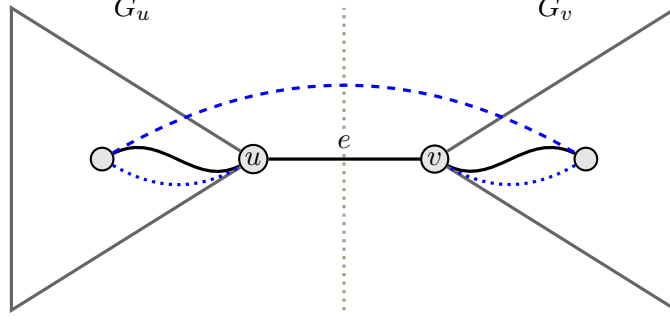


Figure 5: The splitting of z at e . The fractional value of the link crossing the cut (dashed) is added to z^u -value and z^v -value of the left and right shadows (dotted) of the link, respectively.

Figure 5 illustrates the splitting operation. It is easy to see that if y is a feasible fractional solution to the natural LP for the WTAP instance (G, L, c) and $e = \{u, v\} \in E[G]$, then the splitting of y at e produces two feasible fractional solutions for the natural LP: y^u is a feasible solution for G^u and y^v is a feasible solution for G^v . Furthermore, the total weight of y^u and y^v is easy to express in terms of the weight of y and total weight in y of links covering e :

$$c^\top y^u + c^\top y^v = c^\top y + \sum_{\ell \in \text{cov}(e)} c_\ell y_\ell.$$

The latter equality follows from shadow completeness and the fact that for each link $\ell \in L$, y_ℓ is either counted in y^u , or y^v , and it is counted in both if and only if $\ell \in \text{cov}(e)$. Assuming that y also satisfies the thin coverage property (recall that the optimal fractional solution x satisfies this property), then the additive term in the latter expression can also be easily bounded in terms of the maximum cost of any link and the parameter of the thin coverage property:

$$\sum_{\ell \in \text{cov}(e)} c_\ell y_\ell \leq M \cdot y(\text{cov}(e)) \leq \frac{2M}{\epsilon}.$$

Furthermore, clearly y^u and y^v also satisfy the thin coverage property in the corresponding trees G^u and G^v , so the splitting operation does not violate this property.

Next, we use the splitting operation to decompose the tree \bar{G} into simpler trees. We employ a greedy procedure that maintains a set of pairs \mathcal{T} , each comprising a subtree of \bar{G} and a fractional solution of the natural LP for this subtree. We initialize by setting $\mathcal{T} = \{(\bar{G}, x)\}$. At each iteration, the algorithm chooses an arbitrary pair $(T, z) \in \mathcal{T}$ and checks if it contains a *thin edge*, a notion that we define next.

Definition 10. Let T be a subtree of \bar{G} , let $z \in \mathbb{R}_{\geq 0}^L$ and $\alpha \in \mathbb{R}_{\geq 0}$. An edge $e = \{u, v\} \in E[T]$ is called α -thin with respect to z if the total cost of links connecting nodes in $V[G^u]$, and the total cost of links that connect two nodes in $V[G^v]$ is at least α , namely if

$$\sum_{\ell \in L, \ell \in V[G^u] \times V[G^v]} c_\ell z_\ell \geq \alpha \quad \text{for } q = u, v.$$

Formally, the algorithm selects any $\alpha(M, \epsilon)$ -thin edge with respect to z for

$$\alpha(M, \epsilon) = \frac{4M}{\epsilon^2},$$

removes (T, z) from \mathcal{T} , adds to \mathcal{T} the pairs (T^u, z^u) and (T^v, z^v) , obtained from (T, z) by splitting of z at e and proceeds to the next iteration. If no $\alpha(M, \epsilon)$ -thin edge is found, the

algorithm reports (T, z) as part of the final decomposition of (\bar{G}, x) , removes it from \mathcal{T} and proceeds to the next iteration. After at most $|V[\bar{G}]| - 1$ iterations \mathcal{T} is empty, at which stage the algorithm terminates and returns the full decomposition $(T^1, z^1), \dots, (T^k, z^k)$ of (\bar{G}, x) .

The decomposition produced by the latter algorithm has several useful properties, which we state and prove next. First, each pair (T^j, z^j) in the decomposition can be seen as a WTAP instance (T^j, L^j, c) , where $L^j = L \cap (V[T^j] \times V[T^j])$, and for which z^j is a feasible fractional solution of the natural LP (without the bundle constraints). Formally, $z^j \in \mathbb{R}_{\geq 0}^L$, but from the way that the splitting operation works, we have $\text{supp}(z^j) \subseteq L^j$, so we can indeed interpret it as a solution of the instance (T^j, L^j, c) .

Furthermore, the decomposition satisfies the following property, which will be important later, when we exploit the bundle constraints in the bundle LP. Informally speaking, the way weight is shifted around does not decrease the total fractional cost used to cover any subset of edges, provided that this subset is contained in some part of the decomposition.

Lemma 11. *Let $j \in [k]$ and let $F \subseteq E[T^j]$ be any set of edges. Then*

$$\sum_{\ell \in \text{cov}(F)} c_\ell z_\ell^j \geq \sum_{\ell \in \text{cov}(F)} c_\ell x_\ell.$$

Next, we show that the total increase in cost incurred by the decomposition is very small.

Lemma 12. *Let $(T^1, z^1), \dots, (T^k, z^k)$ be a decomposition of (\bar{G}, x) produced by the greedy procedure. Then*

$$\sum_{i \in [k]} c^\top z^i \leq (1 + \epsilon) c^\top x.$$

In addition, the trees in the decomposition have a convenient structure that we define next.

Definition 13. Let $\beta \in \mathbb{Z}_{\geq 1}$. Call a pair (T, z) β -simple if there exists a node $u \in V[T]$ the removal of which results in a forest with trees K_1, \dots, K_t , such that for each $j \in [t]$

- $\sum_{\ell \in L, \ell \in V[K_j] \times V[K_j]} c_\ell z_\ell \leq \beta$, and
- K_j has at most β leaves.

A node u that leads to such a partition is called a β -center of (T, z) .

Figure 6 illustrates the definition of a β -simple pair. Informally speaking, (T, z) is β -simple (for small β) if by removing a single node from T , one can break it into trees, where each tree has a few leaves, as well as low cost on links that are fully contained in the tree. We stress that a tree of a β -simple pair is *not* a tree that can be decomposed into constant size, or constant depth trees (as in [4]) by removing a single node. Instead, the number of leaves in each part is small, while the total number of nodes can be very large.

The following lemma states that every tree-solution pair in the obtained decomposition is β -simple for

$$\beta(M, \epsilon) = \frac{48M}{\epsilon^2}.$$

Lemma 14. *Let $(T^1, z^1), \dots, (T^k, z^k)$ be a decomposition of (\bar{G}, x) produced by the greedy procedure. Then every pair in the decomposition is $\beta(\epsilon, M)$ -simple for $\beta(\epsilon, M) = \frac{48M}{\epsilon^2}$.*

This concludes the description of the decomposition and its properties. However, before we proceed with the second phase of the algorithm, in which each part of the decomposition is rounded to an integral solution, we need to take care of the following small technicality.

While the union of the trees in the decomposition contains all nodes in \bar{G} , it does not contain all edges in \bar{G} . More precisely, the $k-1$ edges used in the $k-1$ splitting operations that resulted in the final decomposition are not part of any tree in the decomposition. Hence, we need to explain how these edges are covered in the solution returned by the algorithm. Here, a trivial solution $L_1 \subseteq L$ containing an arbitrary covering link per edge has a cost of at most $(k-1)M$, which is at most an ϵ -fraction of total fractional cost: Each one of the k solutions z^1, \dots, z^k has, by the property of the decomposition, a fractional cost of at least $\alpha(M, \epsilon) = \frac{4M}{\epsilon^2}$, implying that the total fractional cost is at least $\frac{4Mk}{\epsilon^2}$, so $c(L_1) \leq \epsilon \cdot \sum_{i \in [k]} c^\top z^i$.

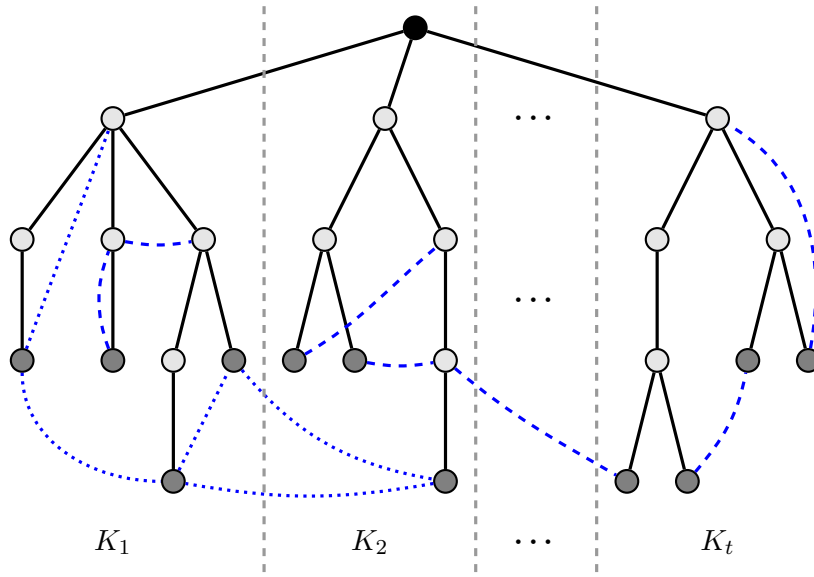


Figure 6: A 4-simple pair (T, z) of an (unweighted) TAP instance. The full node is a 4-center. The shown links represent the support of z , with integral (dashed) and half-integral (dotted) links. The number of leaves, as well as the total fractional weight in each subtree is at most 4.

3.2 Phase 2: Rounding β -Simple Pairs

The second phase of the rounding algorithm accepts a decomposition $(T^1, z^1), \dots, (T^k, z^k)$ of the pair (\bar{G}, x) into $\beta(M, \epsilon)$ -simple pairs, and outputs k WTAP solutions S_1, \dots, S_k , one for each pair. The final output of the algorithm is hence $S^{\text{ALG}} = L_0 \cup L_1 \cup S_1 \cup \dots \cup S_k$, which is feasible for the original WTAP instance. As we have shown, the contribution of $L_0 \cup L_1$ to the cost of the solution is $O(\epsilon)c^\top x$, so its contribution can henceforth be neglected, as it entails an arbitrarily small loss in the approximation guarantee. Furthermore, Lemma 12 guarantees that the total cost of the solutions z^1, \dots, z^k is at most an ϵ -fraction larger than that of x . Finally, since $\text{supp}(z^j) \subseteq V[T^j] \times V[T^j]$ holds for every $j \in [k]$, we can treat each pair (T^j, z^j) as a separate instance-solution pair, thus neglecting the dependencies between the pairs and presenting a rounding procedure for one such pair.

Let (T, z) henceforth denote any pair in the decomposition. We present two rounding procedures, each achieving a good approximation with respect to some part of the fractional solution z . It is then easy to show that the approximation guarantee claimed in Theorem 1 is attained by reporting the solution with the lower cost among the two solutions.

Interestingly, only one of the rounding procedures exploits the bundle constraints in the LP. The other procedure only uses properties of star-shaped solutions and Proposition 8, and provides a good approximation when the instance is *close to being star-shaped*. We present this procedure first. For what remains we fix any $\beta(M, \epsilon)$ -center $r \in V[T]$, call it *root*, and denote by R^1, \dots, R^m the set of trees that are obtained by removing r from T .

3.2.1 First Rounding Procedure: Nearly Star-Shaped Pairs

We call a link $\ell \in V[T] \times V[T]$ a *cross-link* if it connects two nodes in different trees among R^1, \dots, R^m . Observe that any cross-link ℓ has the root r incident to its path P_ℓ^T . A link $\ell \in V[T] \times V[T]$ that is not a cross-link is called an *in-link*. Note that all links ℓ with $r \in \ell$ are in-links.

Define $z^{cr} \in \mathbb{R}_{\geq 0}^L$ and $z^{in} \in \mathbb{R}_{\geq 0}^L$ to be the parts of z that correspond to cross-links and in-links, respectively. Formally, $z_\ell^{cr} = z_\ell$ if ℓ is a cross-link and $z_\ell^{cr} = 0$, otherwise, and $z^{in} = z - z^{cr}$.

The following lemma proves the existence of a simple rounding algorithm that for any $\lambda > 1$ produces a solution with cost at most $\frac{4\lambda}{3(\lambda-1)}$ times the total cost of cross-links in z , albeit at a high cost in terms of in-links.

Lemma 15. *Let $\lambda > 1$ be any constant. Given T and z , there is a polynomial time algorithm that produces a set of links $S \subseteq L$ that covers $E[T]$ with cost at most*

$$c(S) \leq 2\lambda c^\top z^{in} + \frac{4\lambda}{3(\lambda-1)} c^\top z^{cr}.$$

Proof. The rounding algorithm works as follows. Denote by E_λ the set of edges in $E[T]$ that are covered by a fraction of at least $\frac{1}{\lambda}$ by in-links in z . In other words

$$E_\lambda = \left\{ e \in E[T] \mid z^{in}(\text{cov}(e)) \geq \frac{1}{\lambda} \right\}.$$

Let $y = \lambda z^{in}$ be a fractional WTAP solution that covers every edge in E_λ completely (i.e. $y(\text{cov}(e)) \geq 1$ for all $e \in E_\lambda$). Now, invoke Proposition 4 to produce a solution $S^1 \subseteq L$ with cost

$$c(S^1) \leq 2c^\top y = 2\lambda c^\top z^{in}$$

that covers all edges in E_λ . Contract all edges in E_λ to obtain a new tree T' with edge set $E[T'] = E[T] \setminus E_\lambda$.

By definition of E_λ , and from feasibility of z , every edge in $E[T']$ is covered by at least a fraction $\frac{\lambda-1}{\lambda}$ with cross-links, namely

$$z^{cr}(\text{cov}(e)) \geq \frac{\lambda-1}{\lambda}$$

holds for all $e \in E[T']$, so $y' = \frac{\lambda}{\lambda-1} z^{cr}$ is a fractional WTAP solution for the tree T' . Our aim is to cover all edges in T' with cross-links contained in $L^{cr} = \text{supp}(z^{cr})$. Observe, that the WTAP instance restricted to these links is feasible, by definition of T' , and furthermore, it is *star-shaped*, as for any cross-link $\ell \in L^{cr}$, the path $P_\ell^{T'}$ is incident to the root r . We can now apply Proposition 8 to construct a solution $S^2 \subseteq L$ covering all edges of T' with cost

$$c(S^2) \leq \frac{4}{3} c^\top y' = \frac{4\lambda}{3(\lambda-1)} c^\top z^{cr}.$$

Since $S = S^1 \cup S^2$ comprises a feasible solution to the WTAP instance on T , the lemma is proved. Figure 7 illustrates the proof. \square

We apply Lemma 15 to round z with $\lambda = 3 + \sqrt{5}$, whenever $\frac{c^\top z^{cr}}{c^\top z} \geq \alpha^*$, for some $\alpha^* \in (0, 1)$ that we determine later. The values of λ and α^* are simply chosen to minimize the overall approximation guarantee.

We proceed to the second rounding procedure, that is used to round z when the ratio $\frac{c^\top z^{cr}}{c^\top z}$ is small, namely when $\frac{c^\top z^{cr}}{c^\top z} < \alpha^*$.

3.2.2 Second Rounding Procedure: Nearly Decomposable Pairs

Our second rounding procedure proceeds in two steps. First, each cross-link $\ell = uv \in L$ is replaced in the fractional solution z with the two shadows $\ell_u = ur$ and $\ell_v = vr$, by adding z_ℓ to the ℓ_u -th and ℓ_v -th components of z , and setting z_ℓ to zero. This way, we obtain a solution that has no cross-links in its support. Formally, create a new solution $y \in \mathbb{R}_{\geq 0}^L$ derived from z by setting for each $\ell = pq \in L$

$$y_\ell = \begin{cases} z_\ell & \text{if } \ell \in \text{supp}(z^{in}), r \notin \ell \\ z_\ell + \sum_{\ell' \in \text{supp}(z^{cr}), q \in \ell'} z_{\ell'} & \text{if } p = r, q \in V[T] \setminus \{r\} \\ 0 & \text{otherwise.} \end{cases}$$

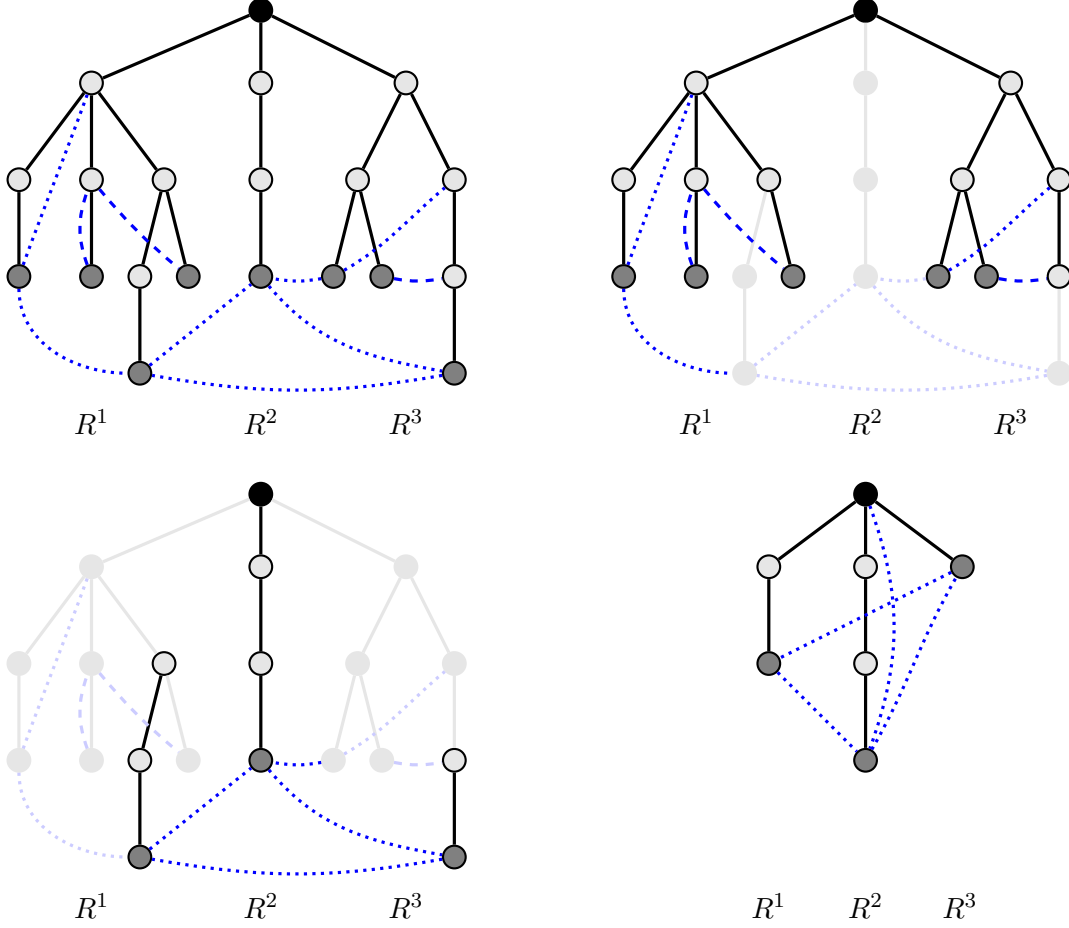


Figure 7: Illustration of the proof of Lemma 15. Top Left: A tree T and the support of z are shown. Dashes and dotted links represent integral and half-integral links, respectively. Top Right: The set E_λ is shown, for $\lambda = 2$, as well as all in-links partially covering E_λ . Bottom Left: The edges of the tree T' and the cross-links partially covering $E[T']$. Bottom Right: The tree T' and the support of the solution y' , obtained by contracting E_λ . The instance is star-shaped.

Clearly, $c^\top y = c^\top z^{in} + 2c^\top z^{cr}$ holds, as the costs of any link is at least the cost of any of its shadows. Next, split the tree T into subtrees $\bar{R}^1, \dots, \bar{R}^m$, by separating T at the root r (note that $\bar{R}^j \neq \bar{R}^j$, as each \bar{R}_j also contains the root r). Now, since y contains no cross-links in its support, it is a union of m disjoint solutions, one for each subtree \bar{R}^j . The latter partition of (T, z) into m parts concludes the first step.

We can henceforth focus on a single part corresponding to a subtree $\bar{R} \in \{\bar{R}^1, \dots, \bar{R}^m\}$, and present a rounding procedure for its corresponding part of y , namely for \bar{y} , defined as

$$\bar{y}_\ell = \begin{cases} y_\ell & \text{if } \ell \in V[\bar{R}] \times V[\bar{R}] \\ 0 & \text{otherwise,} \end{cases}$$

for $\ell \in L$. The union of the obtained solutions comprises a WTAP solution for T . In the following lemma we show how to exploit the bundle constraints to prove that \bar{y} can be rounded up with practically no loss in terms of cost. We charge some of the cost to the early compound nodes and use the fact that (T, z) is β -simple for $\beta = \frac{48M}{\epsilon^2}$.

Lemma 16. *If x is an optimal solution to LP_γ for $\gamma \geq \frac{200M}{\epsilon^2}$, then*

$$\text{OPT}(E[\bar{R}]) \leq c^\top \bar{y} + \sum_{u \in V[\bar{R}]} s_u.$$

Proof. (T, z) is β -simple and the thin coverage property is satisfied thus $c^\top \bar{y} \leq \frac{48M}{\epsilon^2} + \frac{2M}{\epsilon} \leq \frac{50M}{\epsilon^2}$. It follows, from feasibility of \bar{y} and Proposition 4 that $\text{OPT}(E[\bar{R}]) \leq \frac{100M}{\epsilon^2}$. Now, if \bar{R} contains at least $\frac{100M}{\epsilon^2}$ early compound nodes then $\sum_{u \in V[\bar{R}]} s_u \geq \frac{100M}{\epsilon^2} \geq \text{OPT}(E[\bar{R}])$, and we are done.

In the other case the number of early compound nodes in \bar{R} is at most $\frac{100M}{\epsilon^2}$. Our goal is to prove that in this case $E[\bar{R}]$ is a $\frac{200M}{\epsilon^2}$ -bundle in G , and hence if $\gamma \geq \frac{200M}{\epsilon^2}$, LP_γ contains the constraint

$$\sum_{\ell \in \text{cov}(E[\bar{R}])} c_\ell x_\ell \geq \text{OPT}(E[\bar{R}]).$$

Since (T, z) is β -simple, the number of leaves in \bar{R} is at most $\frac{48M}{\epsilon^2} + 1 \leq \frac{50M}{\epsilon^2}$. Let $W \subseteq V[\bar{R}]$ be the set of nodes with degree at least three in \bar{R} . Since the number of leaves is at most $\frac{50M}{\epsilon^2}$, also $|W| \leq \frac{50M}{\epsilon^2}$ holds. Let Q_1, \dots, Q_t be the $t \leq 2 \cdot \frac{50M}{\epsilon^2} = \frac{100M}{\epsilon^2}$ paths obtained by splitting \bar{R} at the nodes in W (see Figure 8).

Now, each path Q_i is a union of paths of G , separated by components contracted in the first phase of the algorithm, when we contracted the edges in E^h , that were covered by L_0 . More precisely, if Q_i is a union of $b \in \mathbb{Z}_{\geq 1}$ paths in G , then these paths are subpaths of Q_i , separated by $b - 1$ early compound nodes. Furthermore, since we designed Q_1, \dots, Q_t to not have nodes of degree larger than two in the interior, each early compound node can lie in the interior of at most one path of the decomposition. Since the total number of early compound nodes is at most $\frac{100M}{\epsilon^2}$, \bar{R} is indeed a union of at most $\frac{200M}{\epsilon^2}$ paths in G , implying that $E[\bar{R}]$ is a $\frac{200M}{\epsilon^2}$ -bundle of G , as required. Figure 8 illustrates this argument.

Consequently, it follows from Lemma 11 and the fact that $E[\bar{R}] \subseteq E[T]$ that

$$c^\top \bar{y} = \sum_{\ell \in \text{cov}(E[\bar{R}])} c_\ell \bar{y}_\ell \geq \sum_{\ell \in \text{cov}(E[\bar{R}])} c_\ell z_\ell \geq \sum_{\ell \in \text{cov}(E[\bar{R}])} c_\ell x_\ell \geq \text{OPT}(E[\bar{R}]),$$

which concludes the proof of the lemma. \square

We have proved that the optimal cost of covering \bar{R} is at most $c^\top \bar{y}$, up to a term that depends on the cost incurred by contracting links of L_0 into early compound nodes that are contained in $V[\bar{R}]$. Now, since $\text{OPT}(\bar{R}) \leq \frac{100M}{\epsilon^2} = O(M)$, we can find an optimal solution by simple enumeration in time $n^{O(M)}$ and report it as a set of links covering \bar{R} (in fact, these values were already computed in the construction of LP_γ). Repeating this for all $\bar{R} \in \{\bar{R}^1, \dots, \bar{R}^m\}$ concludes the rounding procedure for the pair (T, z) . We summarize in the following lemma, which is a direct consequence of Lemma 16 and the previous argument.

Lemma 17. *There is an algorithm that given T and z , computes in time $n^{O(M)}$ a set of links $S \subseteq L$ that covers $E[T]$ with cost at most*

$$c(S) \leq c^\top z^{\text{in}} + 2c^\top z^{\text{cr}} + \sum_{u \in V[T]} s_u.$$

3.3 The Overall Algorithm

The algorithm starts by computing an optimal solution x of LP_γ with $\gamma = \frac{200M}{\epsilon^2}$. For constant M , this LP has polynomial size, and each right hand side of a γ -bundle constraint can be computed using Lemma 18. Then, the first phase of the rounding algorithm computes the set of links $L_0 \cup L_1$ and a decomposition of (\bar{G}, x) into β -simple pairs $(T^1, z^1), \dots, (T^k, z^k)$ for $\beta = \frac{48M}{\epsilon^2}$. This concludes the first phase.

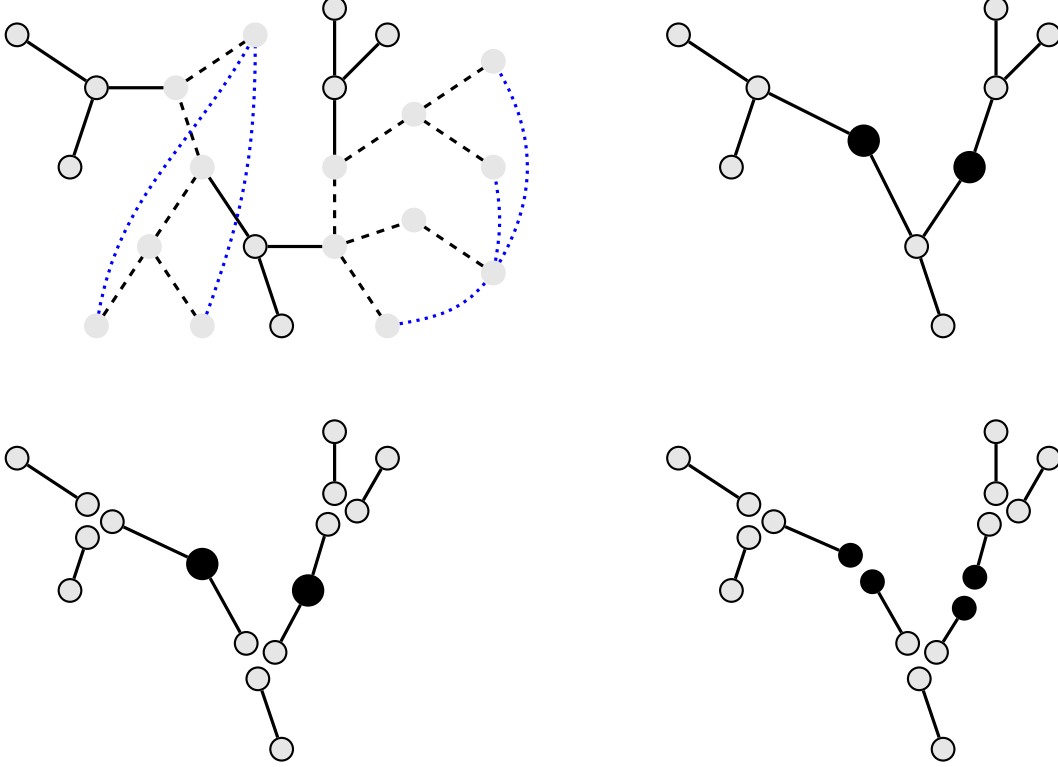


Figure 8: The proof of Lemma 16. Top Left: The part of the original tree G corresponding to \bar{R} . The dashed edges are contracted, as they are part of E^h , resulting in two early compound nodes. The dotted links are included in L_0 . Top Right: The subtree \bar{R} . The large full nodes are the early compound nodes. Bottom Left: Splitting \bar{R} into 7 paths by separating at the nodes in W . Bottom Right: The resulting paths are further subdivided at early compound nodes to obtain 9 paths, forming a 9-bundle in the original tree G .

In the second phase, each pair (T^j, z^j) is rounded using either Lemma 15, or Lemma 17. Concretely, for each $j \in [k]$ the algorithm computes the ratio $\alpha^j = \frac{c^\top z^j{}^{cr}}{c^\top z^j}$ and applies Lemma 15 with $\lambda = 3 + \sqrt{5}$ if $\alpha^j \geq \alpha^*$, and Lemma 17, if $\alpha^j < \alpha^*$, for

$$\alpha^* = 1 - \frac{2(5 - 2\sqrt{5})}{25 + 2\sqrt{5}} \approx 0.96418,$$

which is the value for which Lemmas 15 and 17 give precisely the same approximation, when $\lambda = 3 + \sqrt{5}$. We are ready to prove Theorem 1.

Proof of Theorem 1. Let OPT be the value of the optimal solution. The solution computed by the algorithm is clearly feasible, and the algorithm runs in polynomial time. It remains to prove the approximation guarantee.

The cost of the links in $L_0 \cup L_1$ is at most $2\epsilon c^\top x = O(\epsilon) \cdot \text{OPT}$. Next, for each (T, z) in the decomposition $(T^1, z^1), \dots, (T^k, z^k)$, the cost incurred for the links covering T is at most

$$\min \left\{ 2\lambda c^\top z^{in} + \frac{4\lambda}{3(\lambda - 1)} c^\top z^{cr}, c^\top z^{in} + 2c^\top z^{cr} + \sum_{u \in V[T^j]} s_u \right\} \leq \delta c^\top z + \sum_{u \in V[T^j]} s_u,$$

due to Lemmas 15 and 17. The inequality follows by simple verification, using the definition of the constants λ, α^* and δ and the fact that $z = z^{in} + z^{cr}$. Hence, the total cost of links included

in the solution in rounding all trees in the decomposition is at most

$$\delta \cdot \sum_{i \in [k]} c^\top z^i + \sum_{u \in V[\tilde{G}]} s_u \leq (1 + O(\epsilon)) \delta c^\top x \leq (1 + O(\epsilon)) \delta \cdot \text{OPT},$$

due to Lemma 12 and since $\sum_{u \in V[\tilde{G}]} s_u = c(L_0) \leq \epsilon \cdot \text{OPT}$. This concludes the proof. \square

Acknowledgments The author is grateful to Steve Chestnut and Akaki Mamageishvili for many valuable comments.

References

- [1] J. Cheriyan and Z. Gao. Approximating (unweighted) tree augmentation via lift-and-project, part II. *CoRR*, abs/1507.01309, 2015.
- [2] J. Cheriyan, T. Jordán, and R. Ravi. On 2-coverings and 2-packings of laminar families. In *European Symposium on Algorithms (ESA)*, pages 510–520. Springer, 1999.
- [3] J. Cheriyan, H. Karloff, R. Khandekar, and J. Könemann. On the integrality ratio for tree augmentation. *Operations Research Letters*, 36(4):399–401, 2008.
- [4] N. Cohen and Z. Nutov. A $(1 + \ln 2)$ -approximation algorithm for minimum-cost 2-edge-connectivity augmentation of trees with constant radius. *Theoretical Computer Science*, 489:67–74, 2013.
- [5] G. Even, J. Feldman, G. Kortsarz, and Z. Nutov. A 1.8 approximation algorithm for augmenting edge-connectivity of a graph from 1 to 2. *ACM Transactions on Algorithms (TALG)*, 5(2):21, 2009.
- [6] G. N. Frederickson and J. JáJá. Approximation algorithms for several graph augmentation problems. *SIAM Journal on Computing*, 10(2):270–283, 1981.
- [7] M. X. Goemans, A. V. Goldberg, S. A. Plotkin, D. B. Shmoys, E. Tardos, and D. P. Williamson. Improved approximation algorithms for network design problems. In *SODA*, volume 94, pages 223–232, 1994.
- [8] K. Jain. A factor 2 approximation algorithm for the generalized steiner network problem. *Combinatorica*, 21(1):39–60, 2001.
- [9] S. Khuller. Approximation algorithms for finding highly connected subgraphs. In *Approximation algorithms for NP-hard problems*, pages 236–265. PWS Publishing Co., 1996.
- [10] S. Khuller and R. Thurimella. Approximation algorithms for graph augmentation. *Journal of Algorithms*, 14(2):214–225, 1993.
- [11] G. Kortsarz and Z. Nutov. Approximating minimum cost connectivity problems. In *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2010.
- [12] G. Kortsarz and Z. Nutov. LP-relaxations for tree augmentation. In *Proceedings of the 19th International Workshop on Approximation Algorithms for Combinatorial Optimization (APPROX)*, 2016.
- [13] G. Kortsarz and Z. Nutov. A simplified 1.5-approximation algorithm for augmenting edge-connectivity of a graph from 1 to 2. *ACM Transactions on Algorithms (TALG)*, 12(2):23, 2016.
- [14] Y. Maduel and Z. Nutov. Covering a laminar family by leaf to leaf links. *Discrete Applied Mathematics*, 158(13):1424–1432, 2010.
- [15] H. Nagamochi. An approximation for finding a smallest 2-edge-connected subgraph containing a specified spanning tree. *Discrete Applied Mathematics*, 1(126):83–113, 2003.
- [16] A. Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer Science & Business Media, 2002.

A Solving the Bundle LP

Since the number of γ -bundles is $n^{O(\gamma)}$, it follows that LP_γ can be solved in polynomial time, if for a single bundle $B \in \mathcal{B}_\gamma$, the optimal value $\text{OPT}(B)$ can be computed efficiently.

Consider any γ -bundle B . Contract all edges in $E[G] \setminus B$ to obtain an equivalent instance that corresponds to covering B . Since B is a union of at most γ paths, this instance has at most 2γ leaves. Hence, to compute $\text{OPT}(B)$ it suffices to provide an algorithm that solves an arbitrary WTAP instance with k leaves in time $n^{k^{O(1)}}$, which is the content of the following lemma.

Lemma 18. *Let (G, L, c) be a WTAP instance and let $U \subseteq V[G]$ be the set of leaves in G . Then the WTAP instance can be solved to optimality in time $n^{k^{O(1)}}$, where $k = |U|$.*

Proof. Let $W \subseteq V[G]$ be the nodes in G with degree at least three. Let Q_1, \dots, Q_p be the set of paths obtained by splitting the tree at nodes in W . Since there are k leaves, the number of nodes in W is less than k , and hence $p = O(k)$.

Let $S^* \subseteq L$ be any optimal solution. Now, for each pair of paths Q_i, Q_j , we claim that S^* contains at most two links that connect nodes on Q_i to nodes on Q_j . Indeed, if this is not the case, and there are at least three such links, then necessarily the path of one link is contained in the union of the paths of the other two links, making S^* redundant and contradicting optimality (we assume that L has no links of cost zero, as these links can be included up front in any solution, and the covered edges can be contracted).

Consequently, S^* has at most $O(k^2)$ links connecting nodes on different paths, and all other links in S^* connect nodes that belong to the same path.

The algorithm starts by guessing the $O(k^2)$ links in S^* that connect nodes on different paths. For the correct guess, the problem decomposes into a union of p interval covering problems, which can be solved in polynomial time using dynamic programming. The union of the optimal guess and the optimal solutions from the p paths comprises the reported optimal solution. It is easy to verify the running time. This proves the lemma. \square

B Omitted Proofs

B.1 Proof of Lemma 3

The lemma follows from the fact that the constraint matrix of the natural LP is totally unimodular if all links are up-links, as this implies that there exists an integral solution with the same objective function value, that can be computed by solving the natural LP restricted to the variables corresponding only to up-links.

To prove this claim we use the Ghouila-Houri condition on the rows of the constraint matrix. Consider any subset of constraints, corresponding to a subset $F \subseteq E[G]$ of edges. We show that there exists a subset $F^+ \subseteq F$ such that the constraint obtained by adding all constraints for edges in F^+ and subtracting from the result the sum of the constraints in $F \setminus F^+$ results in a constraint with coefficients in $\{0, 1, -1\}$. This will imply, by the Ghouila-Houri condition, that the constraint matrix is totally unimodular.

First observe that we can assume that $F = E[G]$, as the constraint matrix that corresponds to any subset $F \subseteq E[G]$ is simply the constraint matrix of the natural LP for the tree obtained from G by contracting all edges in $E[G] \setminus F$. To this end define F^+ to be the set of edges that are at an odd distance from r , i.e. the set of edges e , such that the number of edges different from e on the path connecting r to the closest node in e is odd.

It remains to show that for every link $\ell = uv \in L$, it holds that

$$|\{e \in F^+ \mid \ell \in \text{cov}(e)\}| - |\{e \in F \setminus F^+ \mid \ell \in \text{cov}(e)\}| \in \{0, 1, -1\}.$$

This, however, holds because ℓ is an up-link, so the edges in P_ℓ alternate between belonging to F^+ and to $F \setminus F^+$.

B.2 Proof of Lemma 6

The “only if” direction is trivial. To prove the “if” direction consider a solution $S \subseteq L$ that covers all leaf edges. We prove that it is feasible for the WTAP instance. Consider any edge $e \in E[G]$. We show that it is covered by S . Fix any hub $r \in V[G]$. There exists some leaf $u \in V[G]$ of the tree such that e lies on the u - r path in G . Consider the link $\ell \in L$ that covers the leaf edge of u . Since the instance is star-shaped, the path P_ℓ is incident to r , and hence $e \in P_\ell$, as P_ℓ contains all edges on the u - r path in G . It follows that e is covered.

B.3 Proof of Lemma 11

The proof follows easily from the fact that in any splitting operation performed in the algorithm, whenever the fractional assignment of some link ℓ covering some edges in F is decreased to zero, the fractional assignment of a shadow ℓ' of ℓ with the same cost is increased by the same fraction.

B.4 Proof of Lemma 12

Observe that the total number of splittings in the greedy procedure is equal to the number of parts in the final decomposition minus one, namely $k - 1$. Due to the thin coverage property, we know that in each splitting the cost is increased by a total of at most $\frac{2M}{\epsilon}$, so the total increase satisfies

$$\sum_{i \in [k]} c^\top z^i - c^\top x \leq \frac{2kM}{\epsilon}.$$

Now, since we only perform splitting at $\alpha(M, \epsilon)$ -thin edges, each pair (T^j, z^j) in the decomposition satisfies $c^\top z^j \geq \frac{4M}{\epsilon^2}$, hence

$$\sum_{i \in [k]} c^\top z^i \geq \frac{4kM}{\epsilon^2},$$

which implies $\frac{\sum_{i \in [k]} c^\top z^i - c^\top x}{c^\top x} \leq \epsilon$, as desired.

B.5 Proof of Lemma 14

Recall that $x(\text{cov}(e)) \leq \frac{2}{\epsilon}$ for all edges $e \in E[\bar{G}]$. Also, the splitting operation does not increase the coverage level of any edge. More precisely, for any pair (T, z) and any thin edge $e = \{u, v\} \in E[T]$ that are used for splitting in the decomposition algorithm, it holds that $z(\text{cov}(e')) = z^u(\text{cov}(e'))$ and $z(\text{cov}(e'')) = z^v(\text{cov}(e''))$ for any $e' \in E[T^u]$ and $e'' \in E[T^v]$.

Next, recall that pairs (T, z) that comprise the final decomposition have no $\frac{4M}{\epsilon^2}$ -thin edges. This means that for every edge $e \in E[T]$, there exists $u \in e$ such that

$$\sum_{\ell \in L, \ell \in V[T^u] \times V[T^u]} c_\ell z_\ell < \frac{4M}{\epsilon^2}. \quad (5)$$

If there exists an edge $e = \{u, v\}$ for which condition (5) holds for both endpoints u and v , then, by the thin coverage property of the solution we have

$$c^\top z < 2 \cdot \frac{4M}{\epsilon^2} + \sum_{\ell \in \text{cov}(e)} c_\ell z_\ell < \frac{8M}{\epsilon^2} + \frac{2M}{\epsilon} \leq \frac{10M}{\epsilon^2},$$

showing that any node $u \in V[T]$ satisfies the first part of the β' -simple property for $\beta' = \frac{10M}{\epsilon^2}$ in this case. In the other case, every edge $e = \{u, v\}$ has *exactly* one endpoint satisfying (5). Direct each edge $\{u, v\}$ from the node corresponding to the subtree satisfying (5) to the node corresponding to the subtree not satisfying (5). Now, let $u \in V[T]$ be any node with out-degree

zero. By definition of the directed tree, the removal of this node results in subtrees K_1, \dots, K_p satisfying

$$\sum_{\ell \in L, \ell \in V[K_j] \times V[K_j]} c_\ell z_\ell \leq \frac{4M}{\epsilon^2}$$

for all $j \in [p]$, proving existence of a node certifying the first part of the β' -simple property for this case as well.

Until now we proved the existence of a node $u \in V[T]$, the removal of which results in a set of trees K_1, \dots, K_p satisfying the first part of the β' -simple property, for $\beta' = \frac{10M}{\epsilon^2}$. We claim that any such a node is a $\frac{48M}{\epsilon^2}$ -center. Since $\frac{48M}{\epsilon^2} > \frac{10M}{\epsilon^2}$, the first part of the property clearly holds, so it remains to prove that each subtree K_j has at most $\frac{48M}{\epsilon^2}$ leaves. Assume towards contradiction that for some $j \in [p]$, the tree K_j has more than $\frac{48M}{\epsilon^2}$ leaves. Since each link has cost at least 1, and one link can cover at most two leaf edges, the optimal solution to the WTAP instance restricted to the subtree K_j has cost greater than $\frac{1}{2} \cdot \frac{48M}{\epsilon^2} = \frac{24M}{\epsilon^2}$. Consequently, due to Proposition 4, the cost of any fractional solution to the natural LP relaxation on this instance has cost greater than $\frac{1}{2} \cdot \frac{24M}{\epsilon^2} = \frac{12M}{\epsilon^2}$. This leads to a contradicting as follows. Let e' be the edge connecting u to the subtree K_j . Then, z fractionally covers K_j using only the links in $L' = V[K_j] \times V[K_j] \cup \text{cov}(e')$, while

$$\sum_{\ell \in L'} c_\ell z_\ell \leq \frac{10M}{\epsilon^2} + \frac{2M}{\epsilon} \leq \frac{12M}{\epsilon^2},$$

where we used $\sum_{\ell \in L, \ell \in V[K_j] \times V[K_j]} c_\ell z_\ell \leq \frac{10M}{\epsilon^2}$ and $\sum_{\ell \in \text{cov}(e')} c_\ell z_\ell \leq \frac{2M}{\epsilon}$. This concludes the proof of the lemma.

C Better Rounding for TAP (Proof of Theorem 2)

To prove Theorem 2 we only need to prove the following version of Lemma 15. Recall that (T, z) is a pair in the instance decomposition, T is rooted at $r \in V[T]$, which is a β -center of T , and that z^{in} and z^{cr} are the parts of z corresponding to in-links and cross-links, respectively.

Lemma 19. *Given T and z , there is a polynomial time algorithm that produces a set of links $S \subseteq L$ that covers $E[T]$ with cost at most*

$$|S| \leq 2z^{in}(L) + \frac{3}{2}z^{cr}(L).$$

Proof. As in the proof of Lemma 15, we start by replacing each in-link in the support of z with its two shadows. This results in a fractional solution $y \in \mathbb{R}_{\geq 0}^L$ with only cross-links and in-links that are also up-links in the support. We also have

$$y(L) \leq 2z^{in}(L) + z^{cr}(L),$$

as before. The rounding algorithm proceeds as follows. The current tree and the current fractional solutions are denoted by T' and y' , respectively.

- I If some leaf edge of the current tree is covered only by up-links: Include the up-link ℓ covering this leaf edge that covers the most edges, i.e. for which $P_\ell^{T'}$ is the largest. Contract the link ℓ . Go to I.
- II Else, if there exists a link ℓ connecting two leaves $u, v \in V[T']$, choose an arbitrary such link, include it in the solution, and contract all covered edges. Go to I.
- III Else, choose *for every* leaf of T' , one cross-link covering it, and include it in the solution. Return the obtained solution.

We assume that after each contraction operation in the algorithm, all links that become self-loops are removed from the support of y' . We claim that the latter procedure always terminates with a feasible solution S with at most

$$y^{in}(L) + \frac{3}{2}y^{cr}(L) = 2z^{in}(L) + \frac{3}{2}z^{cr}(L)$$

links. We make a few observations. First, notice that links included in step I of the solution do not incur loss in terms of the fractional solution. Indeed, in these iterations a single link is added to the solution, while the total fractional value of the current solution also drops by one unit, since the fractional solution is feasible, and hence the covered leaf edge has at least one unit of fractional links incident to it.

Next, we claim that also in step III no loss is incurred with respect to the fractional solution. Indeed, step III can only be reached if each leaf edge is covered only by fractional up-links and cross-link that connect the corresponding leaf with an internal node in T' . Furthermore, since step I did not materialize, each leaf must have at least one incident such cross-link that is in the support of y' . It follows that, on the one hand, no link in the support of y' covers more than one leaf edge, and on the other hand, it is possible to choose one cross-link incident to each leaf edge. Let q denote the number of leaf edges. The first property implies that $y'(L) \geq q$, as all leaf edges are fractionally covered by y' . The second property implies that there is a set of q cross-links, that cover all leaf edges in T' . Since the instance restricted to cross links is star-shaped (with the root r as its hub), it suffices to choose these q links to cover all of T' . This also proves that the algorithm returns a feasible solution.

It remain to analyze the loss incurred by step II. Since steps I and III incur no loss in terms of the fractional cost, the size of the set S returned by the algorithm is at most

$$y(L) + \sum_{\ell \in \bar{L}} (1 - y_\ell) = y(L) + \sum_{\ell \in \bar{L}} (1 - z_\ell),$$

where \bar{L} is the set of cross-links contracted in step II of the algorithm. Denote $\Delta = \sum_{\ell \in \bar{L}} (1 - z_\ell)$. We show that $\Delta \leq \frac{1}{2}z^{cr}(L)$. This clearly suffices to prove the lemma.

Let $\ell = uv \in \bar{L}$ be a link contracted in step II, where u and v are leaves of T' , and let e_u and e_v be their respective leaf edges. Let y' be the fractional solution in the beginning of that iteration. Let L_u and L_v be the links different from ℓ in the support of y' that cover e_u and e_v , respectively. From feasibility of y' it follows that

$$\sum_{\ell' \in L_p} y'_{\ell'} \geq 1 - y'_\ell$$

holds for $p = u, v$. Furthermore, since $L_u \cap L_v = \emptyset$, it holds that $\sum_{\ell' \in L_u \cup L_v} y'_{\ell'} \geq 2(1 - y'_\ell)$. Define $R = L_u \cup L_v$ and $\theta = 1 - y'_\ell$. We can assume without loss of generality that R contains no in-links. Indeed, every in-link in R is a shadow of ℓ , and hence before contracting ℓ , we can shift weight from any such shadow to ℓ , without increasing the cost of the solution, while maintaining feasibility.

Now, notice that by contracting ℓ the links R become up-links, and are hence never used as contracted links in step II in later iterations. Let $\bar{y} \in \mathbb{R}_{\geq 0}^L$ be the fractional solution in the end of the current iteration. From the latter considerations we have that

$$\bar{y}^{cr}(L) \leq y'^{cr}(L) - \sum_{\ell' \in R} y'_{\ell'} \leq y'^{cr}(L) - 2\theta.$$

Now, since throughout the algorithm we never increase the fraction on any cross-links that we do not immediately contract, it follows immediately from the previous inequality that $2\Delta \leq y^{cr}(L) = z^{cr}(L)$, implying $\Delta \leq \frac{1}{2}z^{cr}(L)$ and proving the lemma. \square

We can now prove Theorem 2 by combining Lemmas 17 and 19.

Proof of Theorem 2. The proof is identical to that of Theorem 1, except that we use Lemma 19 instead of Lemma 15, and we use a different threshold based on the ratio $\frac{z^{cr}(L)}{z(L)}$ (instead of α^*). Concretely, we use Lemma 19 to round the pair (T, z) whenever $\frac{z^{cr}(L)}{z(L)} \geq \frac{2}{3}$, and otherwise we use Lemma 17. The obtained approximation guarantee is not worse than $\frac{5}{3} + \epsilon$ in both cases. \square